THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 275, No. 33, Issue of August 18, pp. 25641–25651, 2000
Printed in U.S.A.

Characterization of Human RhCG and Mouse Rhcg as Novel Nonerythroid Rh Glycoprotein Homologues Predominantly Expressed in Kidney and Testis*

Received for publication, April 19, 2000, and in revised form, May 23, 2000
Published, JBC Papers in Press, June 13, 2000, DOI 10.1074/jbc.M003353200

Zhi Liu‡, Ying Chen‡, Rong Mo§, Chi-chung Hui§, Jan-Fang Cheng¶, Narla Mohandas¶, and Cheng-Han Huang‡‖

From the ‡Laboratory of Biochemistry and Molecular Genetics, Lindsley F. Kimball Research Institute, New York Blood Center, New York, New York 10021, the §Program in Developmental Biology, The Hospital for Sick Children and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5G 1X8, Canada, and the ¶Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

In mammals, the Rh family includes the variable Rh polypeptides and invariant RhAG glycoprotein. These polytopic proteins are confined to the erythroid lineage and are assembled into a multisubunit complex essential for Rh antigen expression and plasma membrane integrity. Here, we report the characterization of RhCG and Rhcg, a pair of novel Rh homologues present in human and mouse nonerythroid tissues. Despite sharing a notable similarity to the erythroid forms, including the 12-transmembrane topological fold, the *RHCG*/*Rhcg* pair is distinct in chromosome location, genomic organization, promoter structure, and tissue-specific expression. *RHCG* and *Rhcg* map at 15q25 of human chromosome 15 and the long arm of mouse chromosome 7, respectively, each having 11 exons and a CpG-rich promoter. Northern blots detected kidney and testis as the major organs of *RHCG* or *Rhcg* expression. *In situ* hybridization revealed strong expression of *Rhcg* in the kidney collecting tubules and testis seminiferous tubules. Confocal imaging of transiently expressed green fluorescence protein fusion proteins localized RhCG exclusively to the plasma membrane, a distribution confirmed by cellular fractionation and Western blot analysis. *In vitro* translation and *ex vivo* expression showed that RhCG carries a complex *N*-glycan, probably at the ^{48}NLS^{50} sequon of exoloop 1. These results pinpoint RhCG and Rhcg as novel polytopic membrane glycoproteins that may function as epithelial transporters maintaining normal homeostatic conditions in kidney and testis.

The Rh antigens, originally identified in human red blood cells (RBCs),[1] are potent immunogens and, when incompatible, cause hemolytic disease of the newborn and blood transfusion reaction (1). They are defined by two erythroid-specific transmembrane (TM) proteins, the Rh polypeptides and the Rh-associated glycoprotein (RhAG), which form a multisubunit complex and display exodomains as D or CcEe antigens (2). The Rh polypeptides are highly polymorphic and are distinguished from RhAG by two biochemical features, *i.e.* palmitoylation but no glycosylation. In contrast, RhAG is largely invariant at the population level and is thought to modulate the assembly of the Rh complex and its surface expression. Human Rh polypeptides and RhAG are encoded by *RHCED* and *RHAG* loci that reside on chromosomes 1 and 6, respectively (3, 4). Despite this difference, these *RH* genes have originated from a common ancestor during evolution in that they are homologous in coding sequence and are grossly similar in exon/intron organization (5, 6).

Although their exact functions remain to be identified, Rh proteins (and their complex) are necessary for the maintenance of RBC morphology and plasma membrane integrity. The Rh deficiency syndrome, a rare inherited form of hemolytic anemia, is caused by mutations at the *RHAG* or *RHCED* locus (2). In this disorder, RBCs deficient in all Rh antigens exhibit spherostomatocytosis and multiple membrane abnormalities (7), implying that Rh proteins have some functional roles in membrane physiological processes. At the level of secondary structure, RhAG and RhCE/D bear a 12-TM topology resembling many transporters (8) and thus are likely to possess a transporter activity across the lipid bilayer. Ammonium or amino group-containing compounds have been suggested as candidate ligands for erythroid Rh proteins, owing to their connective sequence similarity with the NH_4^+ transporters from bacteria, yeast, and plants (2, 9).

The Rh family of genes and proteins is rooted deeply in evolution, and the erythroid homologues of human *RHAG* and *RHCED* genes are coexpressed in RBC of all mammalian species. The closest *RHAG* and *RHCED* relatives are those from nonhuman primates, as they share not only high sequence identity but also some antigen reactivity (10, 11). We previously showed that the mouse erythroid *Rhag* and *Rhced* are identical to their human counterparts in the arrangement of exon/intron junctions, in the conservation of chromosome synteny, and in the pattern of coexpression (12). Furthermore, Rh homologues have been identified in primitive organisms, *e.g.*

* This work was supported in part by National Institutes of Health Grant HL54459 (to C.-H. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF193807, AF193808, AF193809, AF193810, AF183390, AF183391, AF193811, AF193812, AF209468, AF219981 to AF219986, and AF238372 to AF238377.

‖ To whom correspondence should be addressed: Laboratory of Biochemistry and Molecular Genetics, Lindsley F. Kimball Research Inst., New York Blood Center, 310 East 67th St., New York, NY 10021. Tel: 212-570-3388; Fax: 212-570-3251; E-mail: chuang@nybc.org.

[1] The abbreviations used are: RBC, red blood cell; TM, transmembrane; GSP, gene-specific primer; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; GFP, green fluorescence protein; BAC, bacterial artificial chromosome; PAGE, polyacrylamide gel electrophoresis; nt, nucleotide; bp, base pair(s); kb, kilobase(s); UTR, untranslated region; CPMM, canine pancreatic microsomal membranes.

the unicellular slime mold *Dictyostelium discoideum* (2), marine sponge *Geodia cydonium* (13), and earthworm *Caenorhabditis elegans* (14, 15). Structurally, these primitive Rh forms are more similar to RhAG than they are to RhCE/D, providing some important insights into the origin of erythroid members from nonerythroid ancestors. However, it is not known whether these Rh homologues in primitive organisms had branched out, giving rise to new family members expressed in nonerythroid tissues or organs of vertebrate species.

To understand the structure-function relationships of Rh proteins, we have been using a homology approach to clone and identify new Rh genes from lower organisms and from nonerythroid tissues. We describe here the characterization of Rh type C glycoproteins, human RhCG and mouse Rhcg, two orthologous members of the Rh protein family expressed in nonerythroid tissues. We show that RhCG and Rhcg share with their erythroid counterparts the conserved 12-TM topology and cognate signature sequences. However, as the first vertebrate members recognized in the nonerythroid Rh subfamily, RhCG and Rhcg are distinct in primary structure, genomic organization, chromosomal location, tissue-specific expression, and biochemical properties. The studies detailed herein pinpoint RhCG and Rhcg as novel polytopic membrane glycoproteins that may function as epithelial transporters maintaining normal homeostatic conditions in the kidney and testis.

MATERIALS AND METHODS

Cloning of Mouse Rhcg and Human RHCG cDNAs—Using mouse *Rhag* cDNA (12) for a BLAST search (16), a mouse testis expressed sequence tag (AA063867) of high homology was detected. Sense (s) or antisense (a) gene-specific primers (GSPs) were designed to obtain full-length *Rhcg* cDNA. For 5′-RACE (17), 1 μg of mouse testis total RNA was primed with GSP-E7a (GAAGGAATGGACTATCCCTGGCTC). The cDNA was tailed with dCTP using the supplier's protocol (Life Technologies, Inc.) and amplified by two-round PCR with supplied adapter primers and GSPs (E6a1, GGTGGAGAAAATGCCGCAGAA; E6a2, AACCCCACGATGAGAGCGCCGTAA). A 1.1-kb cDNA was purified and sequenced; 3′-RACE (17) was similarly carried out with GSPs (E7s1, CCATTCCTGGAGTCCCGCCTT; E7s2, CGCATCCAGGACACATGTGGCA). To clone *RHCG*, degenerate primers (RhCG-d1, TC(C/T)ATGAC(C/T)ATCCA(C/T)ACATT(C/T)GG; RhCG-d2, CAG(A/G)TT(A/G)TG(A/G)ATGCC(A/G)CATGTGTC) corresponding to conserved *Rhcg* regions (codons 182–189 and 337–344) were used in reverse transcription-PCR of human kidney total RNA. A 480-bp cDNA was purified and sequenced. After comparison with *Rhcg* open reading frame, GSPs were designed and used to derive full-length *RHCG* cDNA, as described above. Other molecular biological procedures followed standard methods (18).

Comparison of RhCG, Rhcg, and Other Homologues—The amino acid sequences of RhCG, Rhcg and other Rh homologues were aligned by means of Clustal W (19) (MegAlign software, DNASTAR). The hydropathy plots were obtained using the Kyte-Doolittle method (20).

RHCG Expression Constructs—The full-length *RHCG* cDNA was cloned in pCR2.1 vector (Invitrogen) using the following GSPs: E10a(*Xho*I), CCGCTCGAGCTAGGGTACCAAGGGTACCGA; E1s(*Bgl*II), GAAGATCTAGCATGGCCTGGAACACCA. All constructs were derived from this master template with *Pfu* DNA polymerase and were verified to be free of mutations by sequencing. To tag the green fluorescence protein (GFP), RHCG was amplified by GSP-E1s(*Bgl*II) plus E10a(*Xho*I) or E1s(*Bgl*II) plus E10a(*Sal*I) and subcloned in the pEGFP-C1 or pEGFP-N3 vector (CLONTECH). For translation and expression studies, *RHCG* was cloned in pYES2 or pcDNA3.1/MycHisA (Invitrogen) by *Bam*HI and *Xho*I double digestion. To express the RhCG C-tail (amino acids 417–479), its coding region was amplified by GSP-E9s(*Bam*HI) (CGGGATCCAGATTACCATTCTGGGGAC) plus E10a (*Xho*I) and cloned separately into vectors pGEX-4T-1 (Amersham Pharmacia Biotech) and pET30a(+) (Novagen).

Production of Polyclonal Antisera for RhCG C-tail—The RhCG C-tail was expressed in *E. coli* BL21 (0.3 mM isopropylthio-β-galactoside at 30 °C for 4 h) as a glutathione *S*-transferase fusion protein or a His_6-tagged peptide. After sonication, glutathione *S*-transferase-RhCG was purified on glutathione Sepharose 4B, and ~300 μg was emulsified with adjuvant and injected into rabbits five times (21). Antisera were affinity-purified by passing through a glutathione *S*-transferase column (Pierce) and then a Ni-NTA column (Qiagen) bound with His-tagged RhCG tail. After washing, the antibodies were eluted with 4 M $MgCl_2$ and dialyzed in 1× phosphate-buffered saline at 4 °C. The IgG fractions showed binding to recombinant RhCG C-tails only.

Screening of Genomic Clones and Definition of Exon/Intron Boundaries—Bacterial artificial chromosome (BAC) clones of *RHCG* or *Rhcg* were screened by PCR. The GSPs for *RHCG* (E6s, CGTGGGTACCGCTGCTGAGAT; E7a, TATGATGCCAGGAATGCCATGCAG) gave a 360-bp band. The GSPs for *Rhcg* (E6s, GCTCTCATCGTGGGGTTCTTCTGC; E7a, CAGGTTGTGRATGCCACATGTGTC) gave a 310-bp band. BAC DNA was fingerprinted by exon PCR and mapped by Southern blotting. Exon/intron boundaries were amplified in two steps from genomic libraries (5, 12) and sequenced. Exon/intron junctions were defined by alignment with cDNA sequences. Introns were amplified from BAC DNA, and their sizes were estimated by gel electrophoresis.

5′-RACE and 5′ Genomic Walking—5′-RACE was performed as mentioned above. The 5′ region of *RHCG* or *Rhcg* gene was amplified from the human or mouse genomic libraries (5, 12) using adaptor primers and exon 1 GSPs. The PCR products were purified and sequenced.

Chromosomal Mapping of RHCG and Rhcg Genes—Fluorescence *in situ* hybridization to interphase chromosomes (22) was used to map human *RHCG*. The DNA from BAC clone 300M8 or 2160I19 was labeled as probes. *Rhcg* was assigned using Jackson BSS interspecific backcross ((C57BL/6JEiXSPRET/Ei)F1XSPRET/Ei) (23). The *Hha*I restriction site was present in *Rhcg* intron 4 of C57BL/6JEi but not in intron 4 of SPRET/Ei. *Rhcg* linkage was defined by *Hha*I digestion of intron 4 fragments in all 94 progenies amplified with GSPs (CTCACAGTGACCTGGATCCTCTAC and CATATCCAACTTGCCCTTCTTGTG).

Northern Blot Analysis and RNA in Situ Hybridization—Two sets of Northern blots (CLONTECH) were hybridized to *RHCG* or *Rhcg* cDNA. The *RHCG* (codon 250–437) and *Rhcg* (codon 194–377) probes were 563 and 552 bp in size, respectively. As a control, the β-actin probe was used.

RNA *in situ* hybridization to mouse embryos and tissues was carried out using standard protocols, as described (24). A 384-bp sequence covering the 3′ region of *Rhcg* cDNA (nucleotides 1110–1494) was subcloned in pCRScript Sk(+) vector (Stratagene) and used to prepare ^{33}P-labeled probes. The antisense and sense probes were generated by *in vitro* transcription of the *Bam*HI- and *Not*I-linearized *Rhcg* plasmids by T7 and T3 RNA polymerases, respectively.

Transfection of RHCG cDNA and Confocal Imaging of RhCG Protein—HEK293 and HeLa cells (ATCC) were grown at 37 °C in Dulbecco's modified Eagle's medium containing 10% fetal calf serum (Life Technologies, Inc.) under 5% CO_2. After 24 h, cells were transferred to either a six-well plate or cover glass and transfected by LipofectAMINE (Life Technologies, Inc.). For transient expression, 1 μg of *RHCG* cDNA was used to transfect 3×10^5 HEK293 or HeLa cells.

For stable selection, *RHCG.myc*/HisA plasmid (1 μg, *Pvu*I cut) was used to transfect 3×10^5 HEK293 cells/well in a six-well plate. Stable cell lines were selected in Dulbecco's modified Eagle's medium (G418, 800 μg/ml) and clonally isolated by standard procedures (18). For confocal imaging, 1 μg of RhCG-GFP or GFP-RhCG plasmid was transfected in 3×10^5 HEK293 or HeLa cells plated onto a 35-mm coverglass (MatTek) and cultured for 24 h. GFP was excited at 488 nm with an argon laser, and the light emitted between 506 and 538 nm was recorded for fluorescein isothiocyanate filter. Images were collected using a Bio-Rad MRC 600 confocal scan head on a Nikon Eclipse 200 microscope with a 60XN.A.1.4 planapo infinity corrected objective. The captured three-dimensional images were processed with Adobe Photoshop (version 4.0).

In Vitro Translation and N-Glycosylation—To define the biochemical features of RhCG, *in vitro* translation and posttranslational processing were carried out. As templates, RhCG plasmids of pYES or pcDNA3.1/MycHisA were used in a transcription-coupled translation system (Promega) with [^{35}S]methionine (15 mCi/ml, Amersham Pharmacia Biotech). Labeled RhCG was analyzed by 12% SDS-PAGE (25) after incubation with or without canine pancreatic microsomal membranes (CPMM) (Promega).

Western Blot and N-Glycanase Analysis of RhCG—Cellular fractionation and isolation of membrane vesicles from stable HEK293 cells were done as described with minor modification (26). Protein was resuspended at 1 mg/ml in ice-cold buffer (10 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 250 mM sucrose). RBC ghosts were prepared as described (27). 8 μg of HEK293 membranes and 50 μg of RBC membranes, treated with or without PNGase F (New England BioLabs), were analyzed by 12% SDS-PAGE (25). Western blots were probed by purified anti-RhCG (1:1,500) or horseradish peroxidase-conjugated anti-Myc monoclonal antibody (1:5,000) (Invitrogen). The anti-RhCG probed blot was stained

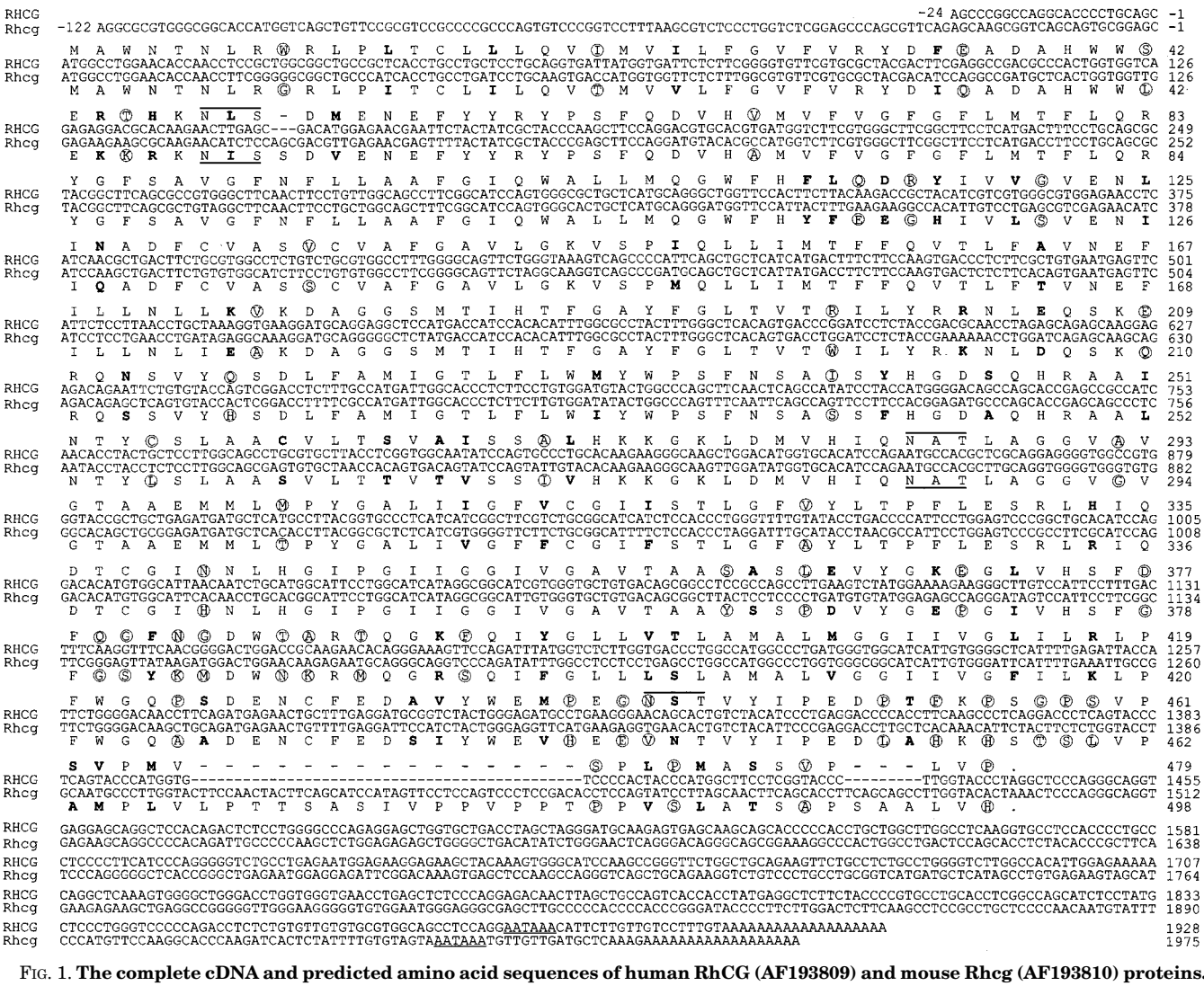

FIG. 1. **The complete cDNA and predicted amino acid sequences of human RhCG (AF193809) and mouse Rhcg (AF193810) proteins.** The sequences are aligned using the Clustal W program. Amino acid changes of similar nature are shown in *boldface,* and those of different nature are *circled*. Deletions are denoted by *dashes*. The *N*-glycosylation motifs in RhCG are *overlined,* and those in Rhcg are *underlined*. Nucleotide and amino acid numbers are counted by reference to the first position of ATG codon as +1. The polyadenylation signal AATAAA in the 3′-UTR of the two genes is *underlined*.

with horseradish peroxidase-linked donkey anti-rabbit IgG (1:5,000) (Amersham Pharmacia Biotech) and visualized by a chemiluminescent kit (Pierce). The anti-Myc probed blot was visualized directly with the above detection system.

RESULTS

Nucleotide Sequences of RHCG and Rhcg cDNAs—Sequencing revealed that *RHCG* and *Rhcg* cDNAs, 1952 and 2097 bp long, respectively, share an overall 68.8% identity at the nucleotide level, forming an orthologous pair. The variation in the sequence is mainly in the 5′-UTR and 3′-end of open reading frames (Fig. 1). *RHCG* and *Rhcg* have a consensus polyadenylation signal but lack the typical Kozak sequence (28) and stop codon preceding the first in-frame AUG codon, two features common to the erythroid Rh genes (5, 12). However, *RHCG* and *Rhcg* are G/C-rich (*RHCG,* 57.5%; *Rhcg,* 55.4%; *versus RHAG,* 42.6%; *Rhag,* 41.8%) and have completely different 5′- and 3′-UTR sequences.

The Primary Structure and Predicted Topology of RhCG and Rhcg Proteins—The open reading frame of human *RHCG* encodes a 53-kDa, 479-amino acid polypeptide, whereas that of mouse *Rhcg* encodes a 55-kDa, 498-amino acid polypeptide (Fig. 1). The RhCG and Rhcg proteins are highly conserved, sharing 77.2% identity and 90.4% similarity with regard to their primary structures. RhCG differs from Rhcg by three small deletions: one serine at position 51 and two amino acid stretches in the extreme C-terminal region are absent in RhCG (Fig. 1). The majority of variations are conserved amino acid changes, with ~50% of them being clustered at the C-terminal positions of 360/361.

The RhCG (pI = 6.2) and Rhcg (pI = 6.5) proteins are negatively charged at physiologic pH and are composed mainly of hydrophobic (44%) and polar (25%) amino acids. Hydropathy analysis (20) showed that they share identical topology and charge distribution, each spanning the lipid bilayer 12 times, with both N and C termini facing the cytoplasm (Fig. 2). The 12-TM fold and its cognate signatures, including the charged residues (Asp^{129} and Glu^{166}), are conserved and reminiscent of other homologs, *i.e.* the RhAG/Rhag (4, 12) and RhBG/Rhbg pairs (Fig. 3).[2] However, RhCG and Rhcg are distinct in that they have a much elongated C-terminal segment and unique N-terminal and exoloop sequences. There are also three *N*-glycosylation N*X*(S/T) motifs in RhCG and two in Rhcg (Fig. 1),

[2] Z. Liu and C.-H. Huang, unpublished observations.

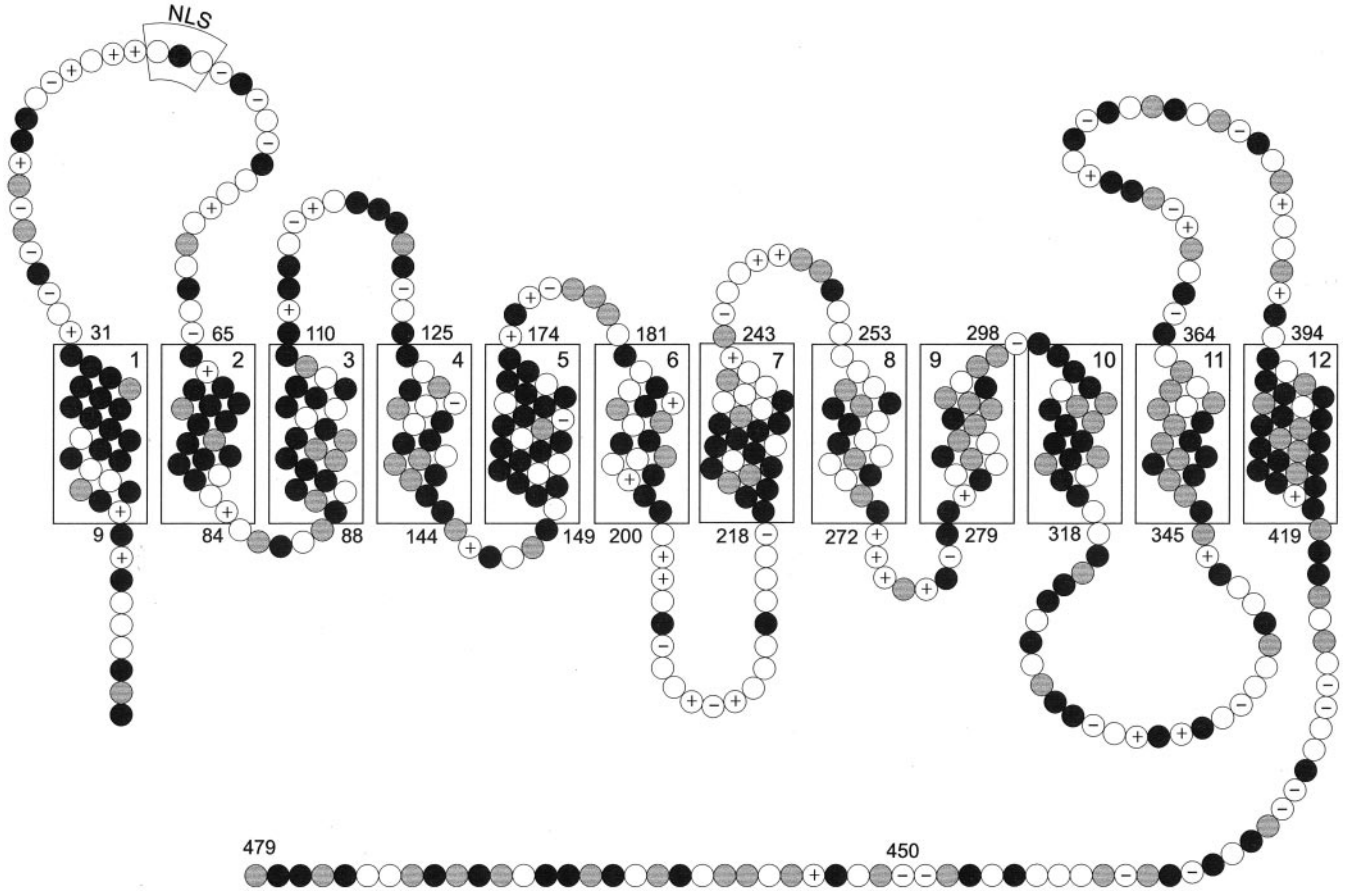

FIG. 2. **The membrane topology and charge distribution of RhCG.** Human RhCG and mouse Rhcg show a nearly identical hydropathy profile as calculated in Kyte-Doolittle scale. For brevity, only the model for RhCG is shown. The 12-TM helices and their amino acid positions defining the boundary are indicated. Different *circles* denote groups of amino acids: *solid circles,* hydrophobic Phe, Ile, Leu, Met, Val, and Trp; *gray circles,* Gly, Ala, and Pro; *open circles,* polar Ser, Cys, Thr, Asn, Gln, and Tyr; +, positively charged Lys, Arg, and His; and –, negatively charged Asp and Glu. The only *N*-glycosylation site present in the first exo loop is illustrated.

suggesting that these proteins are expressed as polytopic glycoproteins. Nevertheless, except for ^{48}NLS^{50} (RhCG) (Fig. 2) or ^{48}NIS^{50} (Rhcg), other N*X*(S/T) motifs were predicted to reside in the TM or in the cytoplasmic domain and thus are unlikely to become glycosylated *in vivo*.

Relationship of RhCG/Rhcg with Other Rh Protein Homologs—Along with the identification of RhCG and Rhcg, we also isolated cDNAs encoding new Rh homologues from model organisms and nonerythroid tissues.[2] Multiple sequence alignment yielded a dendrogram in which the known Rh homologues are clustered in three distinct groups: primitive, erythroid, and nonerythroid (Fig. 3). RhCG and Rhcg were placed as a pair in the nonerythroid subgroup that also includes Hs.RhBG, Mm.Rhbg, Dr.Rhg, and Gc.Rhg. All these proteins are polytopic glycoproteins that are likely to contain a single *N*-glycan on their exoloop 1.

As shown by pairwise sequence comparison, RhCG and Rhcg are most closely related to the nonerythroid RhBG/Rhbg pair (51.7–54.5% identity)[3] and erythroid RhAG/Rhag pair (45.4–50.9% identity) (Fig. 3). This degree of overall identity is largely attributable to conserved TM segments (particularly TM2–11) and their immediately adjacent amino acids, suggesting that these domains define an important function. Significantly, both RhCG and Rhcg are much less similar to antigen carriers, RhCE/D, or Rhced (24.4–28.5% identity). Considering their relationship with the primitive subgroup, RhCG and Rhcg are more closely related to the fruit fly homologue Dm.Rhp, with which they share a 37% overall sequence identity (Fig. 3). These results suggest that the *RHCG/Rhcg* pair originated from early primitive gene precursors and was subjected to an independent evolutionary pathway following its duplication and separation from erythroid members.

Chromosomal Location of RHCG and Rhcg Genes—The chromosomal location of the human *RHCG* gene was mapped by fluorescence *in situ* hybridization using the two BAC genomic clones as probes. The result showed that *RHCG* is located at 15q25 of chromosome 15 (Fig. 4*A*). Using *RHCG* as a query for BLAST search, we detected an exactly matched expressed sequence tag (U75833) from the human CEPH YAC clone that had been mapped at 15q26 (29). These data establish a separate location of *RHCG* from the erythroid *RHCED* and *RHAG* loci that are mapped at chromosomes 1p34–36 and 6p11–21.1, respectively (3, 4).

The mouse *Rhcg* gene was mapped by linkage analysis (Fig. 4*B*). *Rhcg* is nonrecombinant to the *D7Xrf229* locus, with an LOD score of 28.3. This linkage placed *Rhcg* distal to *Gnb2-rs1* but proximal to *Pcsk3* on the long arm of chromosome 7, a region showing conserved synteny with human 15q25 containing *RHCG* gene (Fig. 4*A*). Notably, this locus map is also separate from the erythroid *Rhced* and *Rhag* loci that are localized to chromosomes 4 and 17, respectively (12).

Genomic Organization and Exon/Intron Structures—BAC clones retaining human *RHCG* or mouse *Rhcg* gene were isolated and characterized in order to delineate their structural

[3] Z. Liu and C.-H. Huang, manuscript in preparation.

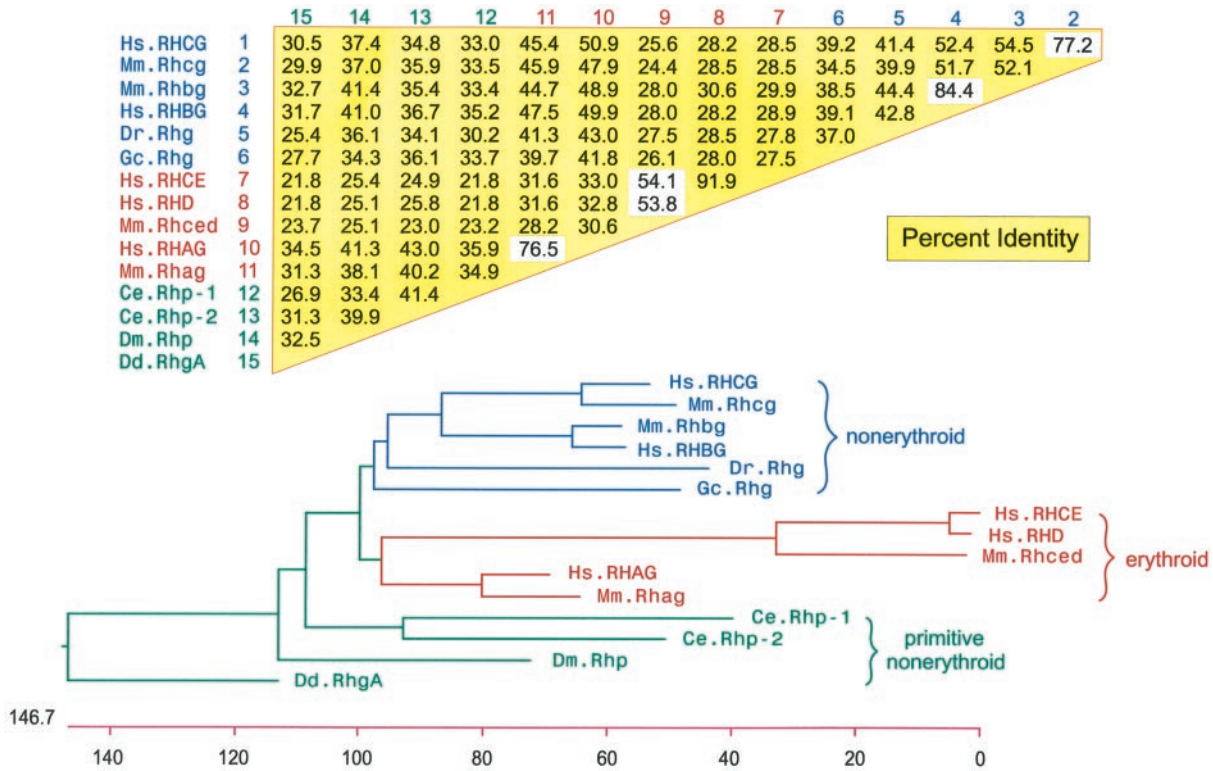

FIG. 3. **Percentage of identity and phylogenetic relationship between RhCG/Rhcg and other homologs.** *Upper panel,* percentage of identity of protein sequences of the various Rh homologs obtained by MegAlign. *Boxed area* shows the percentage of identity of the human and mouse orthologous pairs. *Lower panel,* phylogenetic tree of the Rh family of proteins. Organisms are as follows: *Hs, Homo sapiens*; *Mm, Mus musculus* (house mouse); *Dr, Danio rerio* (zebrafish); *Gc, G. cydonium* (marine sponge); *Ce, C. elegans* (nematode); *Dm, Drosophila melanogaster* (fruit fly); and *Dd, D. discoideum* (slime mold). Note that both RhCG/Rhcg and RhBG/Rhbg belong to the nonerythroid group, as they merge at the same branch point. The GenBank™ accession numbers are as follows: Hs.RHBG, AF193807; Mm.Rhbg, AF193808; Hs.RHCG, AF193809; Mm.Rhcg, AF193810; Hs.RHAG, AF031548; Mm.Rhag, AF057526; Hs.RHD, L08429; Hs.RHCE, X54534; Mm.Rhced, AF057524; Dr.Rhg, AF209468; Gc.Rhg, Y12397; Ce.Rhp-1, AF183390; Ce.Rhp-2, AF183391; Dm.Rhp, AF193812; and Dd.RhgA, AF193811.

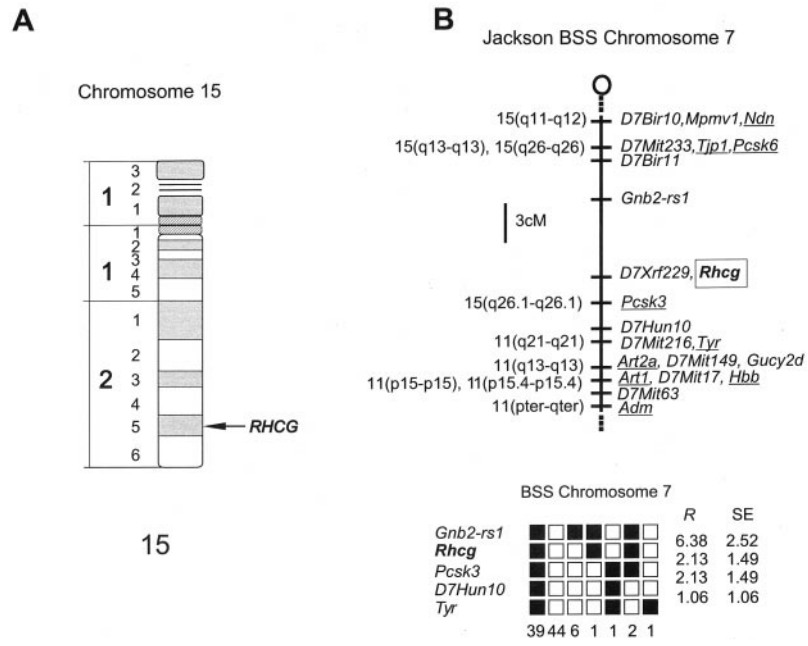

FIG. 4. **Chromosomal location of human *RHCG* and mouse *Rhcg*.** *A,* diagram for the location of *RHCG* at 15q25. DNAs purified from *RHCG* BAC clones 300M8 and 2160I19 were used as fluorescence *in situ* hybridization probes and gave the same result. *B, Rhcg* map on chromosome 7 (the centromere toward the *top*). A 3-cM *scale bar* is indicated. Loci mapping to the same position are listed in alphabetical order. Corresponding human map positions for underlined loci are listed to the *left* of the chromosome bar.

organization. As shown in Fig. 5, the two genes share a nearly identical organization, each having 11 exons that range from 74 to 187 bp and 10 introns that range from 0.2 to 8.8 kb. In both *RHCG* and *Rhcg* genes, exon 11 occurs as a noncoding 3′-UTR segment, a unique feature that distinguishes from all other Rh homologs. Their internal exons 2–9 are identical in size, and their exons 2–6 are arranged in the same fashion as exons 2–6 of *RHAG* and *Rhag* genes (Fig. 5). Because exons 2–6 encode TM2–9 domains in both erythroid and nonerythroid homologs, their conservation in sequence and size implies strict control of the length of the corresponding TM domains and flanking loops.

Fig. 6 shows the sequences of *RHCG* and *Rhcg* genes containing splice sites and exon/intron junctions. All 5′ donor and 3′ acceptor sites conform to the "GT-AG" rule and show consensus to the primate and rodent gene splicing signals (30). Of

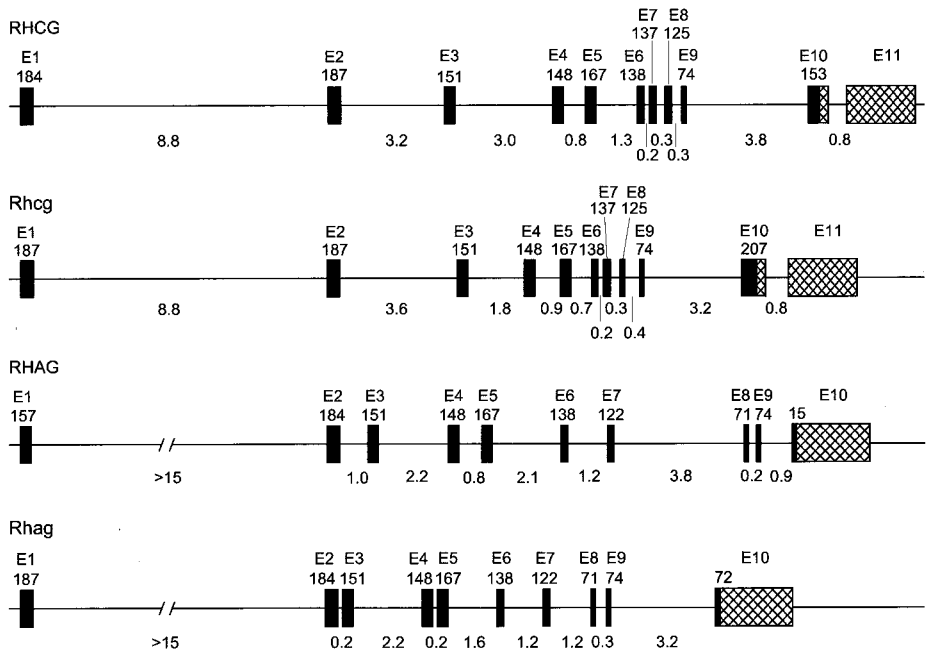

FIG. 5. **Organization of human *RHCG* and mouse *Rhcg* genes and comparison with the erythroid *RHAG* and *Rhag* genes.** The size of exons (E1–E11) is counted in base pairs, and that of introns is counted in kilobase pairs. The noncoding exon for the 3′-UTR downstream of the stop codon in each gene is *shaded*. The human *RHCG* GenBank™ accession numbers are AF219981–AF219986. The human *RHAG* GenBank™ accession numbers are AF238372–AF238377.

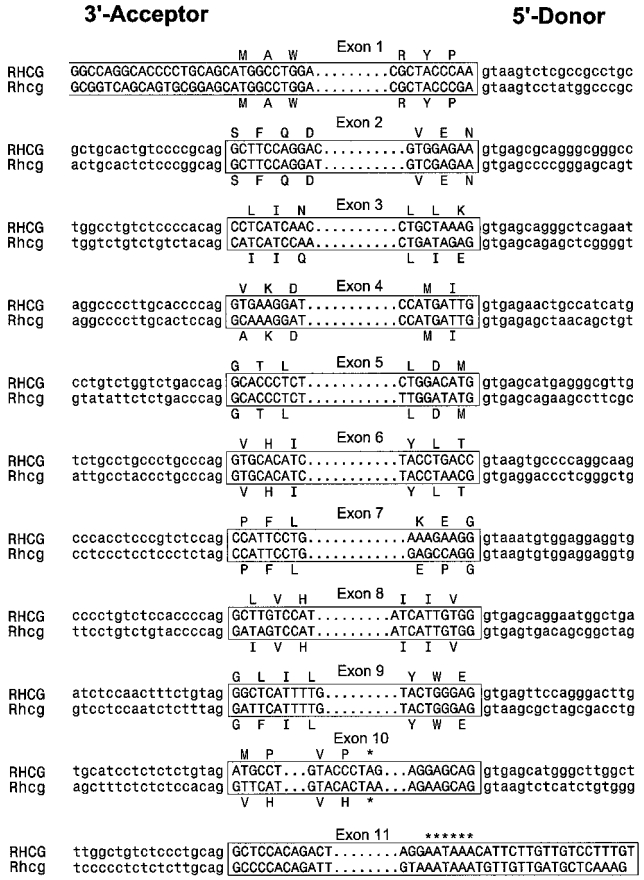

FIG. 6. **Assignment of splice sites and exon/intron junctions between *RHCG* and *Rhcg*.** Exons are in *uppercase letters,* and interval sequences are omitted. Acceptor and donor splice sites are in *lowercase letters*. Amino acids encoded by exon borders are listed. *Stars* denote the stop codon and polyadenylation signal. Note that exon 11 encodes no amino acid sequence.

the 10 coding exons assigned, all but exons 6 and 10 are asymmetrical containing one (exons 1, 3, 4, 5, 7, and 9) or two (exons 2 and 8) split intercodons at the 5′ and/or 3′ exon/intron junction. The amino acids encoded by exon/exon boundaries are the same between RhCG and Rhcg, except for four conserved changes (Fig. 6). Such a division of coding exons is observed in mammals (10–12) but not in more primitive organisms (data not shown).

Proximal Promoter Sequence and Transcription Start Site—Genomic walking revealed that the 5′ region upstream of the first ATG codon of *RHCG* or *Rhcg* contains multiple *cis*-acting elements known to bind various transcription factors (Fig. 7). This proximal sequence is highly enriched in CpG dinucleotides (39 CpGs out of 443 bp for *RHCG* and 25 CpGs out of 279 bp for *Rhcg*), a hallmark of CpG islands that often overlap with generic promoters (31). The sequence is also highly asymmetrical in strand composition alternating with pyrimidine and purine stretches. Notably, these features and many *cis*-acting motifs are absent from erythroid *RH* promoters (5, 6, 12, 32–34).

Sequencing of the 5′-RACE products identified –24A and –126A as the major transcription start sites in the *RHCG* and *Rhcg* genes, respectively (Fig. 7). The sites were determined by using total RNA isolated from the human kidney and mouse testis tissues, and this mapping result was consistent with the beginning of 5′-UTR sequence observed in the full-length cDNAs (Fig. 1). The fact that no in-frame ATG occurs in the genomic sequence further upstream of the transcription start site (Fig. 7) also supports the predicted AUG codon as the translation initiation signal (Fig. 1).

Tissue Expression of RHCG and Rhcg—The expression of *RHCG* and *Rhcg* was assessed by Northern blot analysis (Fig. 8) and by *in situ* hybridization (Fig. 9). In human adult tissues, a 2.0-kb *RHCG* transcript was found abundantly expressed in the kidneys and also in the brain, testis, placenta, pancreas, and prostate (Fig. 8*A*). (Its apparent lower expression in the testis was probably caused by an uneven sampling of testicular tissues for poly(A)+ RNA preparation.) In human fetal tissues, the kidney was the only organ found to express *RHCG* transcripts. In mouse adult tissues, *Rhcg* was shown to be also highly expressed in kidney and testis but not in other tissues, *e.g.* brain (Fig. 8*B*). Furthermore, no *Rhcg* transcript was detected in embryos of 7–19 days of gestation nor in erythroid tissues (data not shown). Taken together, the results indicate that *RHCG* or *Rhcg* is largely expressed at very late stages of development and with limited tissue specificities.

The results from RNA *in situ* hybridization further helped

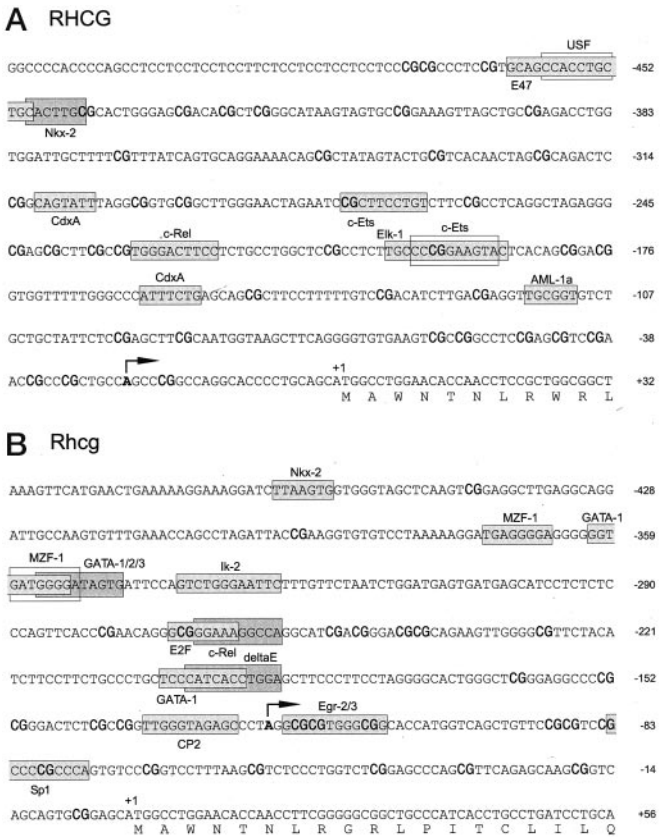

FIG. 7. **Nucleotide sequences of the 5′ region of human *RHCG* and mouse *Rhcg* genes.** *A, RHCG; B, Rhcg*. Potential *cis*-acting motifs known to bind to the various transcription factors are *boxed,* and their orientation is indicated *above* (sense) or *below* (antisense) the sequences. CpG dinucleotides are clustered in the region and are shown in *boldface*. The major transcription start site (–24A in *RHCG* and –122A in *Rhcg*) is denoted by a *bent arrow*. The first position of ATG initiation codon is numbered +1. The partial amino acid sequence of exon 1 is shown.

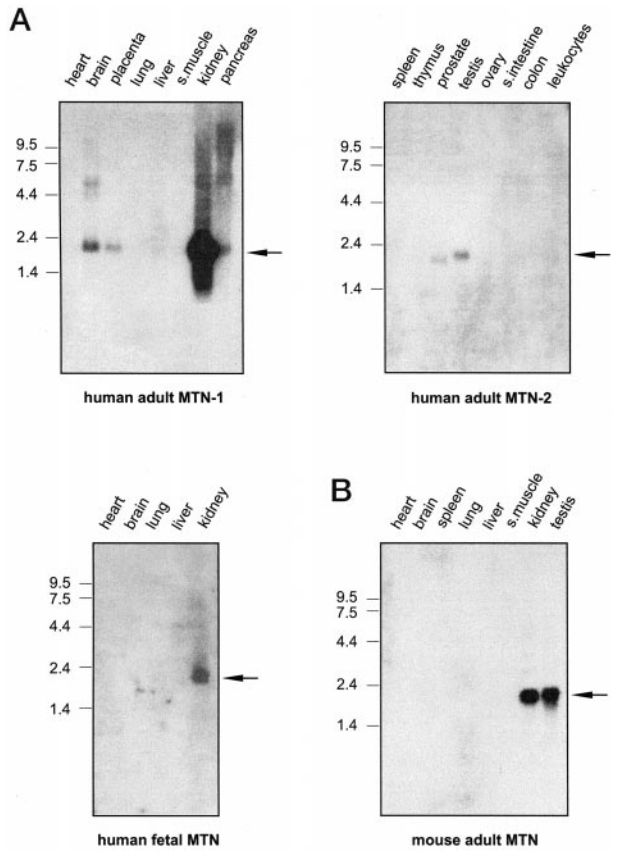

FIG. 8. **Northern blot analysis.** *A, RHCG* expression in human adult and fetal tissues. *B, Rhcg* expression in mouse adult tissues. The size marker is indicated. Tissues are denoted *above* each panel. *s. muscle,* skeletal muscle; *s. intestine,* small intestine. Each lane was loaded with 2 μg of poly(A)+ RNAs. The 2.0-kb major band is indicated with an *arrow*. The actin cDNA hybridization with these blots was relatively uniform (data not shown for brevity).

refining of the sites of *Rhcg* expression in mouse tissues (Fig. 9). In mouse testis, *Rhcg* was abundantly expressed in the seminiferous tubules (Fig. 9*B*), the main components of the adult testicle that produce spermatozoa (35). Higher magnification revealed that the *Rhcg* signals decorated the complex stratified epithelium (Fig. 9*C*), a lining of the seminiferous tubules known to contain spermatogenic cells and supporting Sertoli cells (35). In mouse kidney, *Rhcg* was broadly and abundantly expressed in both the renal cortex and the medulla (Fig. 9*D*). The signals exhibited a radially distributed pattern along their length that corresponded to the parallel alignment of the myriad minute uriniferous tubules that form the functional renal units (35). At higher magnification, the *Rhcg* signal was primarily confined to the epithelial linings of the collecting tubules (Fig. 9, *E* and *F*). These results point toward specific expression of *Rhcg* in the tubular structures of the testis and the kidney.

Cellular Localization of RhCG by Confocal Imaging Analysis—Confocal microscopy (Fig. 10) revealed that the fluorescence signal was confined to the plasma membrane of cells transfected with RhCG-GFP constructs (*B, C, E,* and *F*), whereas it was evenly distributed in the cytoplasm of cells transfected with GFP vector alone (*A* and *D*). The membrane localization was not dependent on the orientation of GFP in the fusion constructs or on the type of cells used in the transfection assay. These results imply that the RhCG protein is normally destined to the plasma membrane.

In Vitro Translation and N-Glycosylation of RhCG Protein—To assess ER translocation and early processing events during biosynthesis, RhCG was translated *in vitro*, in either the absence or presence of CPMM, and analyzed by SDS-PAGE. In the absence of CPMM, *in vitro* translated RhCG, with or without a c-Myc tag, migrated as a single band with a molecular mass less than the predicted molecular mass of 53 kDa (Fig. 11*A*). This mobility anomaly is also seen in erythroid homologues and is most likely the result of the highly hydrophobic nature of these proteins (36). In the presence of CPMM, the RhCG proteins exhibited a slower migration pattern suggesting that they may be differentially glycosylated (Fig. 11*B*). The observed size differences between the CPMM-treated and untreated protein products can be accounted by the size of a single complex *N*-glycan. These *in vitro* studies indicated that RhCG could be glycosylated.

Western Blot and N-Glycanase Analysis of RhCG Protein—To establish the *in vivo* expression of RhCG as a glycoprotein, membrane proteins were isolated from HEK293 cells stably transfected with the *RHCG.myc*/His construct. Western blots probed with the RhCG C-tail polyclonal antiserum showed a major protein band with a molecular mass of ~58 kDa (Fig. 12*A, lanes 3* and *4*); this size was slightly larger than that of *in vitro* glycosylated RhCG (Fig. 11*B, lane 4*). After *N*-glycanase treatment, the size of membrane-bound RhCG was decreased to 46 kDa (Fig. 12*A, lanes 5* and *6*), which was equivalent to that of *in vitro* translated but unglycosylated RhCG (Fig. 11*A, lane 6*). This result indicated that the same AUG initiation codon could be used for RhCG synthesis in both *in vivo* and *in vitro*. Because the stably expressed RhCG pro-

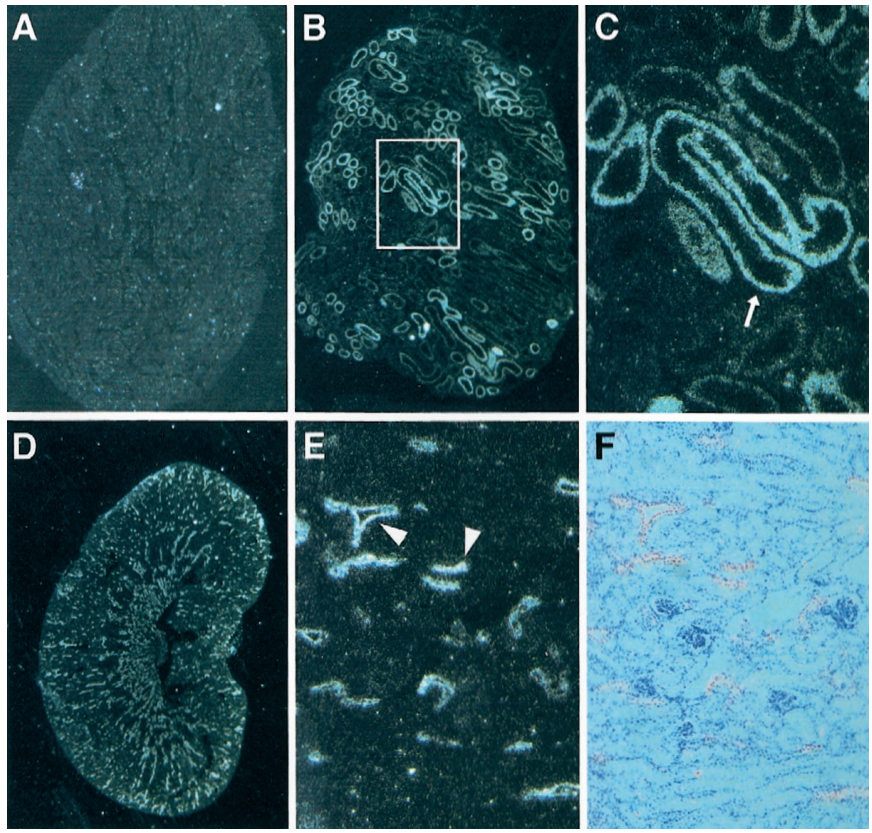

FIG. 9. **Rhcg expression in adult mouse testis and kidney.** *A, in situ* RNA hybridization with a sense control probe. *B,* strong *Rhcg* expression is detected in adult mouse testis. *C,* high magnification of the *boxed area* in *B* showing expression in the seminiferous tubules (indicated by *arrow*). *D,* strong *Rhcg* expression was detected in mouse adult kidney. *E* and *F,* high magnifications showing *Rhcg* expression in the kidney collecting tubules (*arrowheads*). *A–E,* dark field microphotographs; *F,* bright field microphotograph.

tein was expected to carry a Myc/His tag at the C-terminal end, anti-Myc monoclonal antibody was used to probe the same Western blots. This analysis revealed the same banding pattern (Fig. 12*B*), confirming that RhCG exists as a membrane glycoprotein of which the *N*-glycan is most probably attached to the ^{48}NLS^{50} sequon on exoloop 1 (Fig. 2).

DISCUSSION

In the present study, we have isolated several new Rh genes and performed detailed characterization of the human *RHCG* and mouse *Rhcg* genes encoding a novel pair of polytopic plasma membrane glycoproteins. We showed that the RhCG and Rhcg cDNAs are highly conserved with regard to their nucleotide and derived protein sequences and membrane topologic organization. These results firmly establish the genuine orthologous relationship between the two homologues, suggesting that they possess the same functional role(s) in both human and mouse nonerythroid tissues. Clustal analyses of known Rh homologues define RhCG and Rhcg as the first new members of the nonerythroid Rh subfamily in vertebrate species, including all mammals. The detailed studies concerning the genetic and biochemical aspects of RhCG/Rhcg revealed common features as well as distinct differences between the erythroid and nonerythroid Rh homologues. Collectively these results provide novel insights into the molecular evolution, structural conservation, tissue-specific expression, and possible biological function of the Rh family of genes and proteins.

The many homologous sequences assembled suggests the existence of a Rh superfamily, which to date consists of five and four discrete gene members in humans and mice, respectively. Based on their genomic synteny, sequence identity, and tissue specificity, the human and mouse homologues can be divided into four orthologous groups, *i.e.* two erythroid gene pairs (*RHAG*/*Rhag* and *RHCED*/*Rhced*) and two nonerythroid gene pairs (*RHCG*/*Rhcg* and *RHBG*/*Rhbg*). Notably, all the *RH* loci harbor single copy genes, except for human *RHCED* (37), and all have a different map location, although each orthologous group shows comparable chromosomal synteny. By contrast, there is only one Rh gene in the unicellular organism slime mold (2), two tightly linked copies in *C. elegans* (15), and three discrete members in the zebrafish.[2] These observations suggest that both the erythroid and nonerythroid branches came into existence before mammalian radiation and that their multiplicity occurred mainly via intergenic translocation events. This pattern of *RH* duplication is likely to have produced functional novelty in a different temporal and spatial context within the organism, given changes in both coding sequence and regulatory elements of the gene (5–7, 10–12, 32–34).

Compared with erythroid homologues, both primitive and nonerythroid members, including RhCG and Rhcg, are structurally much closer to RhAG/Rhag than they are to RhCED/Rhced. By excluding the more distant RhCED/Rhced group, a membrane fold has been identified, which is characterized by shared signatures conserved among human/mouse RhCG/Rhcg, RhBG/Rhbg, and RhAG/Rhag pairs. This relationship conforms to the approximate equidistance between the erythroid and nonerythroid branches and extends the observation that the *RHCED* series originated from an *Rhag* ancestor gene but evolved at a much increased rate (11, 38). The slower divergence of *RHAG* and the faster evolution of *RHCED* imply an adapted functional specification that pertains to Rh complex assembly in the RBC membrane through heteromeric interactions. These interactions are likely mediated by both N- and C-terminal domains and involve hydrophobic contacts between RhAG and RhCED (39–41). Without RhAG expression, as observed in the regulator type of Rh_{null} disease (40, 41), the RhCE/D proteins and their associated antigens are not dispositioned at the cell surface. In sharp contrast, such heteromeric interaction is not a prerequisite for surface expression of RhCG, as RhCG could by itself reach the plasma membrane, whether transfected into homologous or heterologous cells.

Most likely, nonerythroid Rh homologues possess intrinsic topogenic signal sequences that can direct their own intracellular transport and plasma membrane destination. Further studies are required to determine whether RhCG/Rhcg is on apical or basal membrane in the cells.

Besides their capability of being expressed in heterologous cells, RhCG and Rhcg differ from erythroid members in the following features. 1) They each have a larger exoloop 6 and much longer cytoplasmic C-terminal sequence that is rich in Pro and Ser residues. This novel C-terminal domain may be involved in functional modulation of RhCG or Rhcg through interactions with some yet-to-be recognized intracellular proteins. 2) *RHCG/Rhcg* is unique, having a CpG island-rich promoter and an extra 3′-UTR exon, suggesting a different mode in controlling tissue-specific expression. 3) The human *RHCG* gene is mapped at 15q25 and close to the type I tyrosinemia disease locus (42), although its phenotypic relationship with the latter is unknown. 4) The onset of *Rhcg* expression occurs very late in development, contrary to the early coexpression of *Rhag* and *Rhced* that parallels erythroid differentiation at embryonic stages (12). 5) Compared with the erythroid restriction of *Rhag*/*Rhced* (12), *RHCG* or *Rhcg* is abundantly and broadly expressed in the kidney and in the testis. 6) Although residing in the plasma membrane, as erythroid Rh proteins, Rhcg is concentrated in the epithelial linings of the tubular structures, the kidney collecting tubules, and the testis seminiferous tubules. These features suggest new testable hypotheses as to the function of RhCG/Rhcg, possibly acting as a previously unrec-

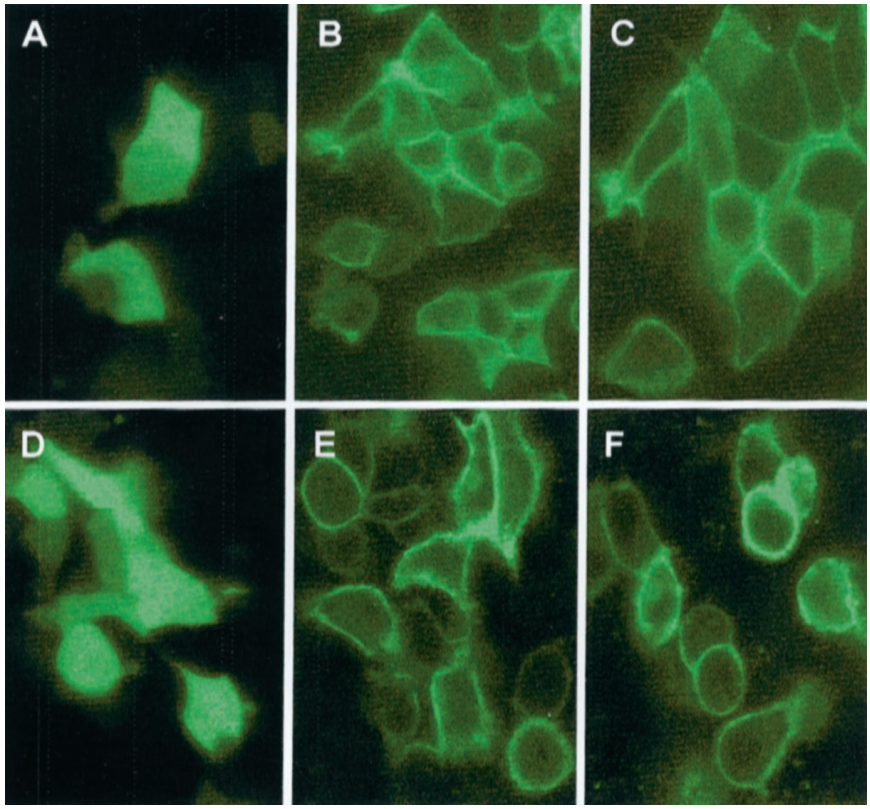

FIG. 10. **Localization of RhCG to the plasma membrane by confocal microscopy.** Cultured HEK293 or HeLa cells were transfected with either *RHCG-GFP* (cloned in pEGFP-N3) or *GFP-RHCG* fusion plasmid (cloned in pEGFP-C1). Images were collected on a Bio-Rad MRC confocal laser scanning microscope. The panels are designated as follows: *A–C,* homologous HEK293 cells; *D–F,* heterologous HeLa cells; *A* and *D,* positive controls (pEGFP-N3 vector alone); *B* and *E*, *RHCG-GFP* fusion construct; *C* and *F, GFP-RHCG* fusion construct.

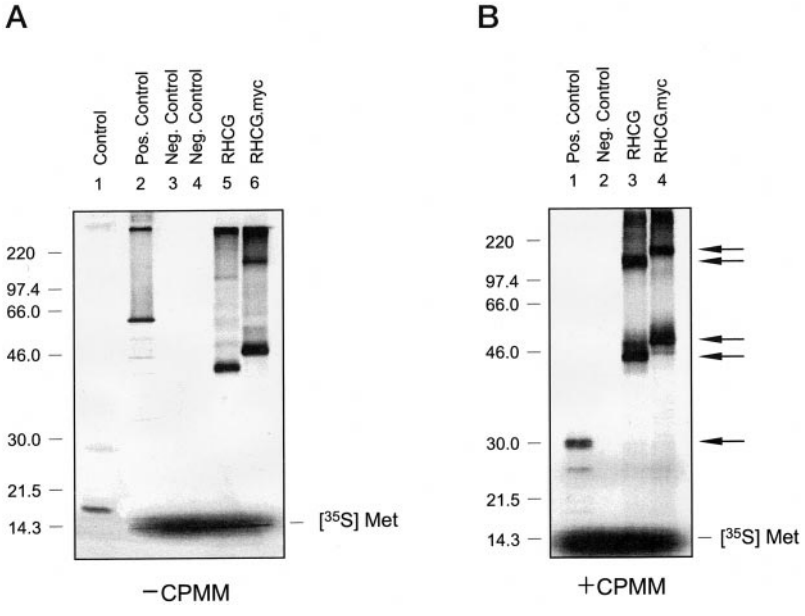

FIG. 11. ***In vitro* transcription-coupled translation and *N*-glycosylation analysis of RhCG protein.** *A,* analysis of *in vitro* translation products by 12% reducing SDS-PAGE in the absence of CPMM. *Lanes 1–4* are positive and negative controls: *lane 1,* yeast α-mating factor; *lane 2,* luciferase; *lane 3,* DNA omitted; *lane 4,* pYES2 vector. *Lane 5,* pYES2/RhCG. *Lane 6,* pcDNA3.1/RhCG-Myc tag. *B,* analysis of *in vitro* translation products by reducing 12% PAGE after incubation with CPMM. *Lane 1,* yeast α-mating factor; *lane 2,* DNA omitted; *lane 3,* pYES/RhCG; *lane 4,* pcDNA3.1/RhCG-Myc tag. Note the up-shift of RhCG in *B*. Molecular mass markers (in kDa) are indicated at the *left margin*.

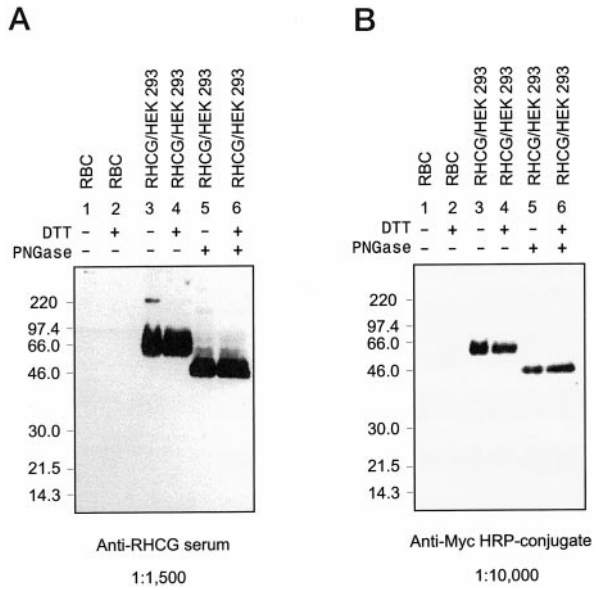

FIG. 12. **Western blot analysis and *N*-glycanase treatment of membrane RhCG.** Plasma membrane proteins were prepared from stable HEK293 cells, fractionated by 12% SDS-PAGE, and blotted onto Hybond NC membranes. RhCG was visualized either by RhCG tail-specific antisera (*A*) or by anti-Myc monoclonal antibody (*B*). Dilution of the two antibodies is shown. Human RBC membranes were used as controls (*lanes 1* and *2*). + and – indicate the presence and absence, respectively, of reducing agent dithiothreitol (*DTT*) and *N*-glycanase PNGase. *HRP,* horseradish peroxidase. Molecular mass markers are shown at the *left margin*.

ognized transporter across the plasma membrane of epithelial cells.

Rh proteins have long been thought to possess a transport activity due to their characteristic TM fold and associated morphological and physiological changes in Rh_{null} disease (2, 7). However, their ligand(s) has not been identified, and their functional dependence on the assembly of a heteromeric complex between RhAG and RhCED remains to be established. Prior studies suggested that Rh might function as an ATP-dependent phosphatidylserine transporter (phosphatidylserine flipase) of the RBC membrane (43, 44). Nonetheless, two lines of evidence disfavor this proposal: 1) Rh proteins lack the ATP-binding cassette (ABC) (2, 7); and 2) Rh_{null} cells have a normal phosphatidylserine transport activity (45).

Recently, a relationship between erythroid Rh proteins and NH_4^+ transporters of the Mep/Amt family has been suggested based on marginal sequence identity (11.1–17.7%) (9). RhCG/Rhcg and other nonerythroid homologues lie in the same range of identity with these transporters. Although NH_4^+ transporters are found in all three domains of life (9), Rh family members occur only in *Eucarya* (2). In *Eucarya*, Mep/Amt members occur in yeast (46), plants (47), and nematodes (14), but not other animals; in contrast, Rh homologues are present in animals (10) and slime mold (2) but absent in yeast and plants, except algae.[2] The sequence connection, along with the tissue specificity, raises the possibility that RhCG and Rhcg occur as potential NH_4^+ transporters in kidney and testis. However, it should be noted that the homology between RhCG/Rhcg and Mep/Amt proteins is quite dispersed and only distantly related. This evolutionary distance could be large enough to endow RhCG/Rhcg with a pivotal structural difference in functional specification. Furthermore, evolution economy also argues against the coexistence of two Rh homologues and three Mep/Amt proteins as the same functional duplicates that all transport NH_4^+ in *C. elegans* (14). Finally, the topology of RhCG/Rhcg resembles numerous 12-TM transporters lacking an ABC domain (8) but differs from that of Mep/Amt members, which generally contain 10 TM helical domains (46–48). If the spatially conserved amino acids are to form a binding site for the amino group only, RhCG or Rhcg may serve to transport some novel amines rather than ammonium itself. The conserved structure and abundant expression suggest that RhCG and Rhcg play an essential role in maintaining normal homeostatic conditions in kidney and testis. In future work, it will thus be important to examine the above hypotheses using models amenable to biochemical and physiological manipulations. To this end, the transient and stable expression systems described here should serve as useful tools for the functional identification of both erythroid and nonerythroid Rh family proteins.

Acknowledgments—We are grateful to Michael Cammer (Director of Imaging Analysis at the Albert Einstein College of Medicine) for confocal microscopy and to Mary Barter (The Jackson Laboratory) for Fig. 4*B*. We appreciate Jianbin Peng's assistance in computer analysis and figure illustrations and staff members of the Laboratory of Microchemistry for DNA sequencing. Thanks are also due to Drs. Olga Blumenfeld and Colvin Redman for helpful comments.

REFERENCES

1. Mollison, P. L., Engelfriet, C. P., and Contreras, M. (1997) *Blood Transfusion in Clinical Medicine,* Blackwell Science, Oxford, England,
2. Huang, C.-H., Liu, Z., and Cheng, G. (2000) *Semin. Hematol.* **34,** 150–165
3. Cherif-Zahar, B., Mattei, M.-G., Le Van Kim, C., Bailly, P., Cartron, J.-P., and Colin, Y. (1991) *Hum. Genet.* **86,** 398–400
4. Ridgwell, K., Spurr, N. K., Laguda, B., MacGeoch, C., Avent, N. D., and Tanner, M. J. A. (1992) *Biochem. J.* **287,** 223–228
5. Huang, C.-H. (1998) *J. Biol. Chem.* **273,** 2207–2213
6. Matassi, G., Chérif-Zahar, B., Raynal, V., Rouger, P., and Cartron, J.-P. (1998) *Genomics* **47,** 286–293
7. Agre, P., and Cartron, J.-P. (1991) *Blood* **78,** 551–563
8. Henderson, P. J. (1993) *Curr. Opin. Cell Biol.* **5,** 708–721
9. Marini, A. M., Urrestarazu, A., Beauwens, R., and Andre, B. (1997) *Trends Biochem. Sci.* **22,** 460–461
10. Blancher, A., and Socha, W. W. (1997) in *Molecular Biology and Evolution of Blood Group and MHC Antigens in Primates* (A. Blancher, J. Klein, W. W. Socha, eds) pp. 147–218, Springer Verlag, Heidelberg
11. Huang, C.-H., Liu, Z., Apoil, P.-A., and Blancher, A. (2000) *J. Mol. Evol.,* in press
12. Liu, Z., Huang, C.-H. (1999) *Biochem. Genet.* **37,** 119–138
13. Seack, J., Pancer, Z., Muller, I. M., and Muller, W. E. (1997) *Immunogenetics* **46,** 493–498
14. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., *et al.* (1994) *Nature* **368,** 32–38
15. Huang, C.-H., Liu, Z., Cheng, G. J., and Chen, Y. (1998) *Blood* **92,** 1776–1784
16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
17. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 8998–9002
18. Sambrook, J., Fristch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
20. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
21. Harlow, E., and Lane, D. (1988) *Antibodies: A Laboratory Manual* 1st Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
22. Heng, H. H. Q., Squire, J., and Tsui, L.-C. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 9509–9513
23. Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J., and Birkenmeier, E. H. (1994) *Mamm. Genome* **5,** 253–274
24. Hui, C.-c., and Joyner, A. L. (1993) *Nat. Genet.* **3,** 241–246
25. Laemmli, U. K. (1970) *Nature* **227,** 680–685
26. Fong, A. D., Handlogten, M. E., and Kilberg, M. S. (1990) *Biochim. Biophys. Acta* **1022,** 325–332
27. Huang, C.-H., Blumenfeld, O. O., Reid, M. E., Chen, Y., Daniels, G. L., and Smart, E. (1997) *Blood* **90,** 391–397
28. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148
29. Chumakov, I. M., Rigault, P., Le Gall, I., Bellanne-Chantelot, C., Billault, A., Guillou, S., Soularue, P., Guasconi, G., Poullier, E., Gros, I., *et al.* (1995) *Nature* **377** (suppl.), 175–297
30. Shapiro, M. B., and Senapathy, P. (1987) *Nucleic Acids Res.* **15,** 7155–7174
31. Gardiner-Garden, M., and Frommer, M. (1987) *J. Mol. Biol.* **196,** 261–282
32. Cherif-Zahar, B., Le Van Kim, C., Rouillac, C., Raynal, V., Cartron, J.-P., and Colin, Y. (1994) *Genomics* **19,** 68–74
33. Huang, C.-H. (1996) *Blood* **88,** 2326–2333
34. Iwamoto, S., Omi, T., Yamasaki, M., Okuda, H., Kawano, M., and Kajii, E. (1998) *Biochem. Biophys. Res. Commun.* **243,** 233–240
35. Fawcett, D. W., and Raviola, E. (1994) *A Text Book of Histology* 12th Ed., Chapman & Hall, New York
36. Moore, S., Woodrow, C. F., and McClelland, D. B. L. (1982) *Nature* **295,** 529–531

37. Carritt, B., Kemp, T. J., and Poulter, M. (1997) *Hum. Mol. Genet.* **6,** 843–850
38. Kitano, T., Sumiyama, K., Shiroishi, T., and Saitou, N. (1998) *Biochem. Biophys. Res. Commun.* **249,** 78–85
39. Ridgwell, K., Eyers, S. A., Mawby, W. J., Anstee, D. J., and Tanner, M. J. (1994) *J. Biol. Chem.* **269,** 6410–6416
40. Huang, C.-H., Cheng, G.-J., Liu, Z., Chen, Y., Reid, M. E., Halverson, G., and Okubo, Y. (1999) *Am. J. Hematol.* **62,** 25–32
41. Cherif-Zahar, B., Raynal, V., Gane, P., Mattei, M.-G., Bailly, P., Gibbs, B., Colin, Y., and Cartron, J.-P. (1996) *Nat. Genet.* **12,** 168–173
42. McKusick, V. A. (1999) in *The Human Genome Database Project*, Johns Hopkins University, Baltimore, MD, Online Mendelian Inheritance in Man
43. Connor, J., and Schroit, A. J. (1988) *Biochemistry* **27,** 848–851
44. Schroit, A. J., Bloy, C., Conner, J., and Cartron, J.-P. (1990) *Biochemistry* **29,** 10303–10306
45. Smith, R. E., and Dalek, D. L. (1990) *Blood* **76,** 1021–1027
46. Marini, A. M., Soussi-Boudekou, S., Vissers, S., and Andre, B. (1997) *Mol. Cell. Biol.* **17,** 4282–4293
47. Kaiser, B. N., Finnegan, P. M., Tyerman, S. D., Whitehead, L. F., Bergersen, F. J., Day, D. A., and Udvardi, M. K. (1998) *Science* **281,** 1202–1206
48. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996) *Science* **273,** 1058–1073